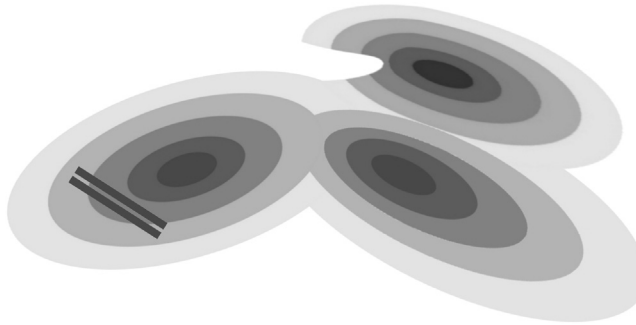
CHAPTER 8

Composite structures

Mateusz Banach[1], **Leszek Konieczny**[2], **Irena Roterman**[1]
[1]Department of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Krakow, Poland
[2]Chair of Medical Biochemistry, Jagiellonian University – Medical College, Krakow, Poland

Contents



Conceptual diagram illustrating a large, complex protein with several roles (p-p interaction, ligand binding cavity), along with fragments characterized by linear distribution of hydrophobicity. In such a protein the active site may be constructed from multiple structural units, each of which – when considered on its own – follows the micellar pattern.

Analysis of a nonredundant set of PDB proteins reveals that a vast majority of individual domains exhibit strong micelle-like characteristics [1]. A

From Globular Proteins to Amyloids
ISBN: 978-0-08-102981-7
https://doi.org/10.1016/B978-0-08-102981-7.00011-7

large protein comprising several chains which are further subdivided into domains, may be regarded as a composite structure, where each unit contributes some of the required information. Multiple units may cooperate to form an active site characterized by specific deviations from the monocentric distribution of hydrophobicity. In certain cases, such active sites are not only discordant but in active opposition to the monocentric distribution — for example, they may exhibit a linear pattern of alternating bands of high and low hydrophobicity.

Single-chian enzyme consisting of three domains

The first example of a composite structure is provided by a single-chain enzyme: hydrolase (E.C. 3.1.5.1) elongation factor Ef-tu complexed with a GTP analog in the antibiotic pulvomycin [2] (structure available as 2C78 in PDB).

The protein catalyzes the following reaction: dGTP + H_2O => deoxyguanosine + triphosphate.It consists of three domains, with catalytic residues located in domain 1.

Analysis of the molecule as a whole

The structure as a whole is not globular and does not contain a prominent hydrophobic core (Table 8.1). This particular protein interacts with two ligands:

GNP — Phosphoaminophosphonic acid-Guanylate ester.

PUL — *(1s,2s,3e,5e,7e,10s,11s,12s)-12-[(2r,4e,6e,8z,10r,12e, 14e,16z, 18s,19z)-10,18-Dihydroxy-12,16,19-Trimethyl- 11,22-Dioxooxacyclodocosa-4, 6,8,12,14,16,19-Heptaen-2- Yl]-2,11-Dihydroxy-1,10-Dimethyl-9-Oxotrideca-3,5,7- Trien-1-Yl 6-Deoxy-2,4-Di-O-Methyl-Beta-L- Galactopyranoside [Pulvomycin]*

Elimination of ligand binding residues lowers the value of RD, which indicates that the binding pocket contains information (although note that even under these conditions RD remains greater than 0.5, suggesting that other factors are at play in terms of producing deviations between O and T).

Elimination of catalytic residues produces a further decrease in RD, which, again, shows that these residues contain information required in the process of catalysis. When considering these residues we also take into account their immediate neighborhood (5 aa in each direction). Notably, 85H is the only catalytic residue whose neighborhood fragment significantly diverges from the theoretical distribution.

Table 8.1 Values of fuzzy oil drop parameters calculated for the structure of elongation factor Ef-tu (2C78) and its three domains. Values listed as boldfaces correspond to fragments of status discordant versus the idealized one; values underlined – fragments representing highly ordered status in respect to idealized distribution.

		RD		Correlation coefficient		
Hydrolase (2C78)	**Fragment**	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
	9–405	**0.627**	**0.593**	0.211	0.385	0.773
Lig		**0.746**	**0.502**	**−0.665**	**−0.005**	**0.647**
No lig		**0.591**	**0.584**	0.260	0.463	0.788
No cat res	21D, 85H	**0.618**	**0.584**	**0.210**	**0.385**	**0.772**
Cat 21D	16–26	0.464	0.464	0.201	0.606	0.786
Cat 85H	79–90	0.465	0.465	0.234	0.490	0.670
Domain 1	9–213	**0.594**	**0.571**	**0.294**	**0.497**	**0.766**
Domain 2	216–308	**0.612**	**0.587**	**0.092**	**0.254**	**0.800**
Domain 3	311–404	**0.637**	**0.569**	**0.120**	**0.311**	**0.775**
Secondary structure						
β-strands	10–17	0.198	0.279	0.336	0.849	0.647
	66–72	0.496	0.213	0.383	0.845	0.722
	75–82	0.228	0.358	0.520	0.586	0.844
	102–108	**0.979**	**0.753**	0.507	0.758	0.533
	130–136	0.626	0.534	0.522	0.309	0.652
	169–173	0.361	0.433	0.446	0.648	0.842
β-sheet		**0.505**	**0.554**	0.342	0.559	0.691
Helices	23–38	**0.698**	0.427	**−0.142**	**0.042**	**0.777**
	46–51	0.395	0.295	0.500	0.648	0.826
	53–60	0.369	0.129	−0.200	0.773	0.371
	85–88	0.388	0.225	0.769	0.597	0.965
	89–97	0.362	0.292	0.402	0.913	0.586
	113–126	**0.556**	**0.516**	**0.303**	**0.433**	**0.881**
	137–141	**0.687**	0.214	**−0.587**	**−0.269**	**0.789**
	143–161	0.643	0.660	0.308	0.408	0.887
	174–185	0.234	0.299	0.453	0.817	0.722
	193–210	0.463	0.397	0.535	0.592	0.924
β-strand	**223–225**	**0.679**	**0.662**	**−0.992**	**−0.974**	**0.940**
	227–231	**0.832**	**0.764**	**−0.489**	**−0.335**	**0.898**
	235–241	**0.886**	**0.984**	**−0.545**	**−0.584**	**0.938**
	246–248	0.175	0.008	0.801	0.900	0.982
	252–256	0.144	0.154	0.743	0.916	0.884
	263–272	**0.603**	**0.716**	**0.257**	**0.013**	**0.822**
	275–277	**0.574**	0.140	**−0.668**	**−0.224**	**0.876**
	279–281	0.169	0.348	0.984	0.997	0.994
	285–290	**0.837**	**0.952**	**−0.749**	**−0.789**	**0.851**

(*Continued*)

Table 8.1 Values of fuzzy oil drop parameters calculated for the structure of elongation factor Ef-tu (2C78) and its three domains. Values listed as boldfaces correspond to fragments of status discordant versus the idealized one; values underlined – fragments representing highly ordered status in respect to idealized distribution.—cont'd

		RD		Correlation coefficient		
Hydrolase (2C78)	**Fragment**	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
β sheet		0.664	0.560	0.660	0.143	0.800
β-strands	311–322	0.238	0.088	0.672	0.812	0.832
	341–343	0.588	0.627	0.528	−0.001	0.848
	347–354	**0.787**	**0.905**	**−0.222**	**−0.338**	**0.871**
	367–380	0.393	0.206	0.387	0.611	0.847
	385–389	**0.529**	**0.945**	**−0.101**	**−0.058**	**0.965**
	393–403	0.248	0.132	0.235	0.807	0.604
β-sheet		0.573	0.422	0.218	0.394	0.777
Helix	324–327	0.448	0.034	0.712	0.587	0.018
Domain A1 – individual						
	9–213	0.394	0.378	0.395	0.695	0.767
Lig. binding		**0.652**	0.303	**−0.228**	**0.221**	**0.648**
No lig. bind.		0.352	0.399	0.487	0.750	0.784
No E		0.390	0.372	0.400	0.700	0.765
Cat E21	61–26	0.553	0.222	−0.187	0.176	0.766
Cat 85H	79–90	0.350	0.218	0.224	0.785	0.669
Secondary structure						
β-strand	10–17	0.255	0.350	0.328	0.853	0.647
	66–72	0.263	0.201	0.187	0.772	0.722
	75–82	0.307	0.247	0.776	0.747	0.843
	102–108	0.946	0.536	0.759	0.795	0.528
	130–136	0.424	0.335	0.132	0.574	0.652
	169–173	0.253	0.317	0.739	0.937	0.842
β-sheet		0.284	0.326	0.463	0.791	0.691
Helices	**23–38**	**0.672**	**0.397**	**0.052**	**0.191**	**0.777**
	46–51	0.426	0.322	0.687	0.558	0.825
	53–60	0.445	0.170	0.070	0.668	0.368
	85–88	0.034	0.016	0.895	0.979	0.966
	89–97	0.224	0.173	0.435	0.894	0.585
	113–126	0.517	0.642	0.262	0.420	0.886
	137–141	**0.707**	0.231	**−0.660**	**−0.347**	**0.789**
	143–161	0.412	0.429	0.411	0.626	0.886
	174–185	0.263	0.330	0.428	0.814	0.720
	193–210	0.391	0.328	0.495	0.623	0.924

Table 8.1 Values of fuzzy oil drop parameters calculated for the structure of elongation factor Ef-tu (2C78) and its three domains. Values listed as boldfaces correspond to fragments of status discordant versus the idealized one; values underlined – fragments representing highly ordered status in respect to idealized distribution.—cont'd

		RD		Correlation coefficient		
Hydrolase (2C78)	**Fragment**	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
Domain A2 – individual						
	216–307	0.310	0.296	0.515	0.770	0.801
Lig. binding		0.309	0.307	0.158	0.753	0.710
No lig. bind		0.310	0.296	0.562	0.772	0.820
Secondary structure						
β-strand	223–225	0.178	0.258	0.684	0.928	0.906
	227–231	0.434	0.333	0.377	0.527	0.897
	235–241	0.234	0.734	0.633	0.834	0.937
	246–248	0.069	0.003	0.990	0.998	0.981
	252–256	0.320	0.306	0.723	0.722	0.957
	263–272	0.363	0.484	0.398	0.743	0.822
	275–277	0.195	0.028	0.925	0.992	0.875
	279–281	0.453	0.678	0.300	0.407	0.993
	285–290	0.170	0.446	0.781	0.880	0.851
β-sheet		0.302	0.217	0.502	0.770	0.799
Domain A3 – individual						
	311–404	0.340	0.298	0.504	0.748	0.767
Lig. binding		0.071	0.117	0.605	0.976	0.635
No ligand		0.342	0.288	0.499	0.751	0.783
Secondary structure						
β-strands	311–322	0.434	0.236	0.447	0.719	0.784
	341–343	0.052	0.061	0.854	0.999	0.849
	347–354	0.314	0.529	0.613	0.756	0.841
	367–380	0.437	0.311	0.473	0.698	0.893
	385–389	0.083	0.489	0.880	0.973	0.954
	393–403	0.286	0.155	0.307	0.793	0.609
β-sheet		0.367	0.277	0.499	0.736	0.771
Helix	324–327	0.272	0.022	−0.321	0.818	−0.348

Going back to the molecule as a whole, we may note that none of its constituent domains are consistent with the theoretical distribution (when analyzed as components of the protein). The status of secondary and supersecondary folds varies; however, all β-sheets are strongly discordant.

Domains treated as independent structural units

In contrast to the above, when considered as standalone units, all three domains exhibit strong accordance with the monocentric core pattern. We will attempt to provide an explanation for this phenomenon.

The final structure, which – for environmental reasons (among others) – must exhibit a nonstandard distribution of hydrophobicity, is formed by joining together three individual components, each of which has been shaped by hydrophobic interactions and resembles a micelle. Even domain 1, which contains the protein's catalytic residues, is regarded as accordant (note that catalytic residues are expected to carry information and therefore exhibit deviations from the Gaussian distribution of hydrophobicity).

The structure also includes ligands, one of which acts as an inhibitor. From the point of view of domains 2 and 3, the status of ligand-binding residues remains highly accordant with the theoretical distribution. This suggests that a structure capable of recognizing and binding the correct ligand emerges as a result of interactions between several independent domains.

The status of catalytic residues (found only in domain 1) varies depending on whether we consider the domain by itself or the molecule as a whole. While a specific deviation has been produced inside the domain for residue H85 (±5 aa), residue E21 does not disrupt its overall micellar pattern. Its status changes (to discordant) only after the respective domains assemble into a composite protein.

The status of catalytic residues can be determined by calculating FOD parameters for the input chain following elimination of such residues. The value of RD decreases in the process, however, it remains greater than 0.5.

Plotting hydrophobicity distribution charts for the entire chain (Fig. 8.1) reveals multiple deviations, particularly in the central area where hydrophobicity is much lower than expected. The presented chart also shows interdomain boundaries. This situation changes when domains are analyzed as standalone units. In particular, Fig. 8.2 reveals strong accordance between T and O for each domain.

The presented analysis suggests that catalytically active structures may be produced by assembling simple micellar components whose O is aligned

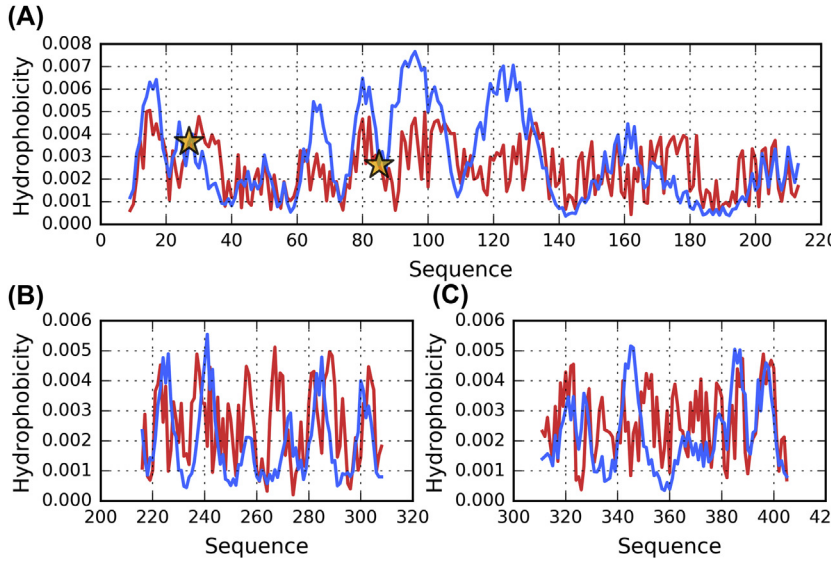


Fig. 8.1 Theoretical (T, blue) and observed (O, red) hydrophobicity distribution profiles for three domains of 2C78 treated as whole protein. (A) domain 1 (9—213), (B) domain 2 (216—307), (C) domain 3 (311—404) Orange stars mark positions of catalytic residues.

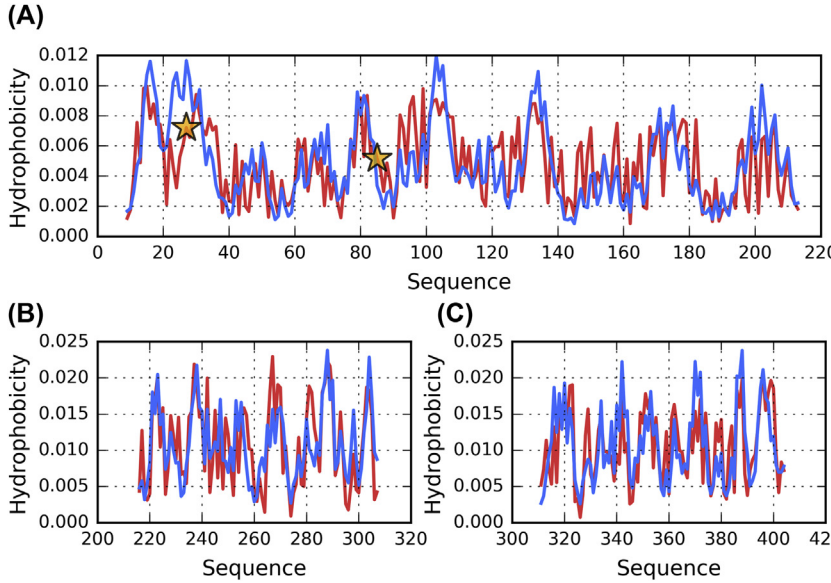


Fig. 8.2 Theoretical (T, blue) and observed (O, red) hydrophobicity distribution profiles for three domains of 2C78 treated as individual units. (A) domain 1 (9—213), (B) domain 2 (216—307), (C) domain 3 (311—404) Orange stars mark positions of catalytic residues.

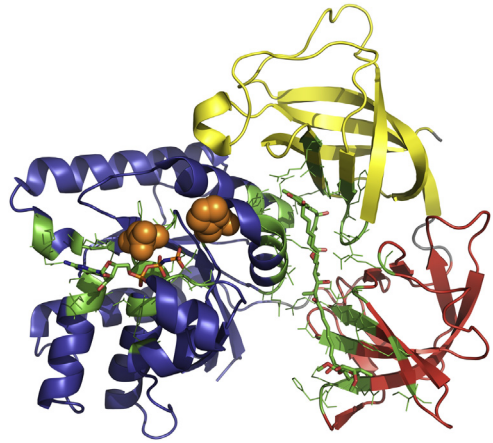


Fig. 8.3 3D presentation of 2C78, with colors distinguishing its three domains: (A) 1 – blue, (B) 2 – red, (C) 3 – yellow. Residues engaged in protein-ligand interactions are green-colored and have side chains displayed. Orange spheres mark catalytic residues (21D, 85H).

with T. The assembly process by itself introduces specific deviations from the monocentric distribution of hydrophobicity, facilitating biological function (Fig. 8.3).

Homodimer enzyme

The second sample protein analyzed in this chapter is the isozyme of citrate synthase from *Sulfolbus tokodaii* strain 7 (PDB ID: 1VGM) [3]. This enzyme is classified as E.C. 2.3.3.1 and participates in the following reaction:

citrate (Si)-synthase – Acetyl-CoA + H_2O + oxaloacetate ⇒ citrate + CoA

The protein is a homodimer (CATH classification: 5.1.1.1), with complex secondary and supersecondary characteristics (Table 8.2). Each monomeric unit is composed of two domains. Structural assessment indicates the lack of a monocentric hydrophobic core (Table 8.2, Figs. 8.4 and 8.5). Elimination of interface residues results in only a slight decrease in RD. Thus, the interface is not thought to significantly disrupt the distribution of hydrophobicity in the dimer. Similarly, elimination of catalytic residues (as well as their immediate neighborhood – 5 aa in each direction) does not produce a marked change in the status of the molecule. Regarding the catalytic residues themselves, their status is highly variable. Residue 189S (together with its neighborhood) conforms to the theoretical distribution,

Table 8.2 Values of fuzzy oil drop parameters calculated for the structure of 1VGM and its two domans. Values listed in boldface indicate conditions which contrast with a micelle-like conformation. Double entries in the top section of the table correspond to chains A and B respectively. "No arms" indicates chains A and B following elimination of "arms" which protrude from their globular structures (fragments 3–13 and 356–378 respectively).

		RD		Correlation coefficient		
Isozyme of citrate synthase (1VGM)	**Fragment**	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
Dimer						
	2 × (3–378)	**0.622**	**0.553**	0.236	0.556	0.736
P-P		**0.604**	**0.531**	0.093	0.600	0.640
No P-P		**0.615**	**0.550**	0.291	0.551	0.760
No E	Eliminated: 189S, 219H 259H, 314D	**0.619**	**0.553**	0.240	0.556	0.739
No E	Enzymatic residues ± 5aa	**0.611**	**0.553**	0.277	0.564	0.758
189S	184–194	0.210	0.214	0.190	0.955	0.405
		0.200	0.195	0.177	0.944	0.409
219H	214–224	**0.579**	0.176	0.095	0.099	0.733
		0.580	0.185	0.101	0.086	0.727
259H	254–264	**0.528**	0.361	0.392	0.243	0.768
		0.353	0.277	0.532	0.630	0.681
314D	309–319	**0.823**	**0.700**	**−0.617**	**−0.343**	**0.642**
		0.805	**0.700**	**−0.672**	**−0.484**	**0.697**
B-sheet	19-23, 28–32, 34-35	**0.761**	0.488	0.283	0.196	0.613
		0.768	0.479	0.328	0.289	0.607
β-sheet	A 12–22 B 359–369	**0.564**	**0.804**	**0.046**	**0.179**	**0.924**
C-term	368–378	0.494	0.296	0.195	0.290	0.771
		0.436	0.225	0.311	0.547	0.794

(*Continued*)

Table 8.2 Values of fuzzy oil drop parameters calculated for the structure of 1VGM and its two domans. Values listed in boldface indicate conditions which contrast with a micelle-like conformation. Double entries in the top section of the table correspond to chains A and B respectively. "No arms" indicates chains A and B following elimination of "arms" which protrude from their globular structures (fragments 3−13 and 356−378 respectively).—cont'd

		RD		Correlation coefficient		
Isozyme of citrate synthase (1VGM)	**Fragment**	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
Chain A						
	3−378	**0.550**	0.480	0.253	0.518	0.720
P-P		**0.571**	0.488	0.030	0.454	0.527
No P-P		**0.553**	0.484	0.305	0.519	0.772
189S	184−194	**0.516**	0.494	−0.323	0.008	0.319
219H	214−224	0.438	0.150	0.447	0.610	0.538
259H	254−264	0.447	0.295	0.483	0.453	0.768
314D	309−319	**0.813**	**0.685**	−0.550	−0.271	0.623
β-sheet	19-23, 28−32, 34−35	**0.684**	0.436	0.238	−0.105	0.500
Chain B						
	3−378	**0.534**	0.465	0.260	0.528	0.715
P-P		**0.558**	0.472	0.005	0.454	0.531
No P-P		**0.524**	0.468	0.320	0.532	0.764
189S	184−194	**0.513**	0.493	−0.316	0.077	0.308
219H	214−224	0.444	0.159	0.506	0.572	0.544
259H	254−264	0.300	0.232	0.564	0.740	0.681
314D	309−319	**0.785**	**0.672**	**−0.607**	**−0.446**	**0.696**
β-sheet	19-23, 28−32, 34-35	**0.628**	0.378	0.277	0.103	0.492
Chain A − no arms						
	14−355	**0.584**	**0.502**	0.284	0.453	0.746
P-P		**0.665**	0.487	0.001	0.184	0.498

No P-P		**0.566**	0.499	0.315	0.471	0.780
189S	184–194	**0.575**	0.444	0.294	0.638	0.286
219H	214–224	0.295	0.087	0.200	0.785	0.536
259H	254–264	0.401	0.258	0.530	0.583	0.768
314D	309–319	**0.822**	**0.700**	**-0.524**	**-0.197**	**0.623**
β-sheet	19-23, 28–32, 34-35	**0.743**	**0.560**	0.204	-0.055	0.414
Chain (B) no arms						
	14–355	**0.582**	**0.501**	0.286	0.460	0.738
P-P		**0.678**	**0.504**	-0.066	0.160	0.510
No P-P		**0.562**	0.495	0.324	0.481	0.770
189S	184–194	**0.654**	**0.635**	**-0.402**	**-0.544**	**0.307**
219H	214–224	0.303	0.093	0.219	0.784	0.546
259H	254–264	0.237	0.180	0.578	0.813	0.682
314D	309–319	**0.805**	**0.700**	**-0.574**	**-0.367**	**0.697**
β-sheet	19-23, 28–32, 34-35	**0.734**	**0.576**	0.305	0.166	0.493
Domain A1						
	15-220, 325-358	0.490	0.433	0.329	0.569	0.689
189S	184–194	0.483	0.459	-0.104	0.292	0.304
219H	214–224	0.333	0.093	0.268	0.786	0.577
β-sheet	19-23, 28–32, 34-35	**0.730**	0.463	0.249	0.106	0.452
Domain A2						
	221–324	0.470	0.339	0.335	0.527	0.820
259H	254–264	0.359	0.324	0.536	0.891	0.733
314D	309–319	**0.758**	**0.609**	0.122	-0.192	0.697

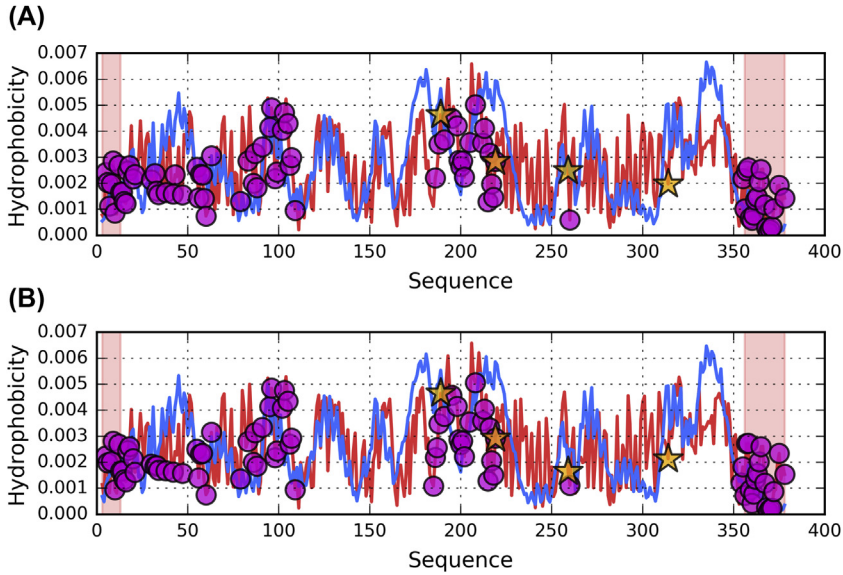


Fig. 8.4 Theoretical (T, blue) and observed (O, red) hydrophobicity distribution profiles for 1VGM. (A) chain A, (B) chain B Magenta circles mark residues engaged in P-P interaction, while stars correspond to catalytic residues. Red background distinguish the location of "arms" protruding from the molecule (fragments 3—13 and 356—378).

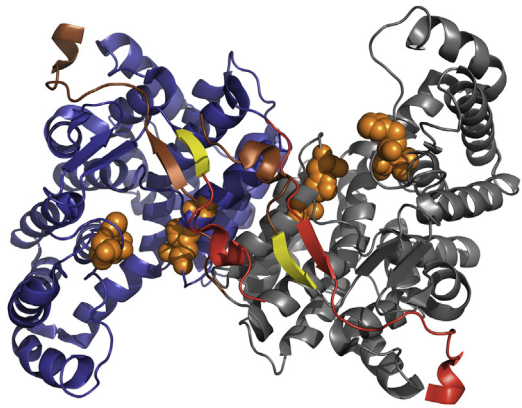


Fig. 8.5 3D presentation of isozyme of citrate synthase (1VGM) homodimer (chain A — blue, chain B — gray). Red fragments in chain A and brown in chain B are "arms" with which the subunits embrace. Yellow fragments (which do not belong to the "arms" themselves) form β-sheets with strands from the "arms". Orange spheres — catalytic residues (189S, 219H, 259H, 314D).

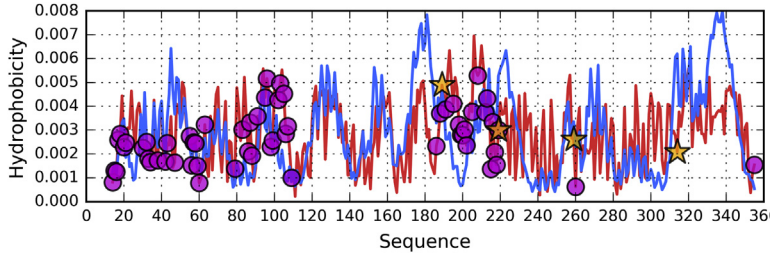


Fig. 8.6 Theoretical (T, blue) and observed (O, red) hydrophobicity distribution profiles for 1VGM chain A without its "arms" – fragment 314–355 (following elimination of fragments 3–13 and 356–378). Magenta circles mark residues engaged in P-P interaction, while stars correspond to catalytic residues.

while residues 219H and 259H diverge from it. 314D, a catalytic residue, actively opposes the theoretical distribution: the conformation of its neighborhood is determined by intrinsic hydrophobicity. Such conditions may have a trickle-down effect on the properties of the protein's aqueous environment.

Strong discordance is also observed for the protein's β-sheets – again, their purpose may be to cause specific structural changes in the solvent.

An interesting structural motif is present in the dimer: two short N-terminal (3–13) and two long C-terminal (356–378) sections ("arms") (Fig. 8.6) protrude from each chian and appear to embrace the other chain. This two-chain structure also includes two antiparallel discordant β-sheets which consist of fragments contributed by both chains: residues 12–22 (chain A) linked with residues 359–369 (chain B) – and *vice versa*.

It should be noted that these fragments – along with the previously mentioned β-sheet – are all found at the "entrance" to the active site (catalytic residues S189, H219, H259, D314). Consequently, they may be suspected of altering the structure of the aqueous environment in order to facilitate the protein's biological function.

Analysis of chains and domains regarded as standalone structures

As already mentioned, the structure of each individual chain consists of a large globular portion and an elongated arm which forms a peculiar "interface". Given that the protruding fragment may significantly hamper FOD analysis of the molecule, it has been excluded from calculations.

Results shown in Table 8.2 reveal an increase in discordance following excision of the protruding "arms". This shows that information is encoded

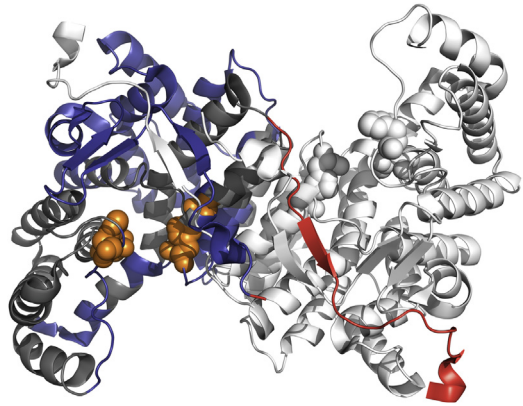


Fig. 8.7 3D presentation of 1VGM homodimer highlighting fragments which conform to a micellar distribution (14—79, 89—100, 109—174, 188—190, 203—205, 211—224, 229—235 and 250—260) — blue. They are shown only in chain A for readability (whole chain B is shown in white). Like on Fig. 8.5, "arm" of chain A is colored red and its four catalytic residues are marked by the orange spheres.

in the globular portion itself. However 'arms" by themselves represent status accordant with expected hydrophobicity distribution. The effect is particularly pronounced for the β-sheet and the catalytic residue at 189S, along with its immediate neighborhood.

Given the complex structure of the enzyme, it is interesting to seek fragments responsible for maintaining its micelle-like form. Strongly micellar fragments include those at 14—79, 89—100, 109—174, 188—190, 203—205, 211—224, 229—235 and 250—260. Taken together, their RD(T-O-R) and (T-O-H) values are 0.454 and 0.348 respectively, with correlation coefficients calculated as 0.352, 0.589 and 0.757 (HvT, TvO and HvO respectively — see Fig. 8.7). They provide the structural backbone which enables the protein to remain stable. Accordance between T and O is evident in Fig. 8.4, particularly for the 109—174 fragment. Note that in order to reduce RD we have eliminated those residues which exhibits major differences between T and O.

Domains

As already remarked, each domain considered on its own (i.e. constructing a 3D Gaussian capsule specifically for that domain) is seen as consistent with the theoretical distribution of hydrophobicity (Table 8.2).

Doman 1, when stripped of ligand-binding residues, reveals a decrease in RD, which indicates that these residues carry information. In some cases,

HvT correlation coefficients calculated for ligand-binding residues adopt negative values, suggesting very strong discordance.

Domain 2 shows excellent accordance with the monocentric distribution and is only slightly disrupted by the presence of the ligand. This accordance is observed both in the neighborhood of ligand-binding residues and in other fragments comprising the domain.

The status of each catalytic residue (along with its 5 aa neighborhood) is as follows:

189S: accordant in the dimer but discordant in the monomer and in the "no arms" structure

219H: discordant in the dimer but accordant in the monomer (including in the "no arms" structure) and in its individual domain

259H: depends on the chain (differing status in chains A and B forming the dimer) but accordant in each individual chain and domain

314D: discordant in all structures subjected to analysis

The above results show that domains, folding individually, shape the neighborhoods of residues 189S, 219H and 259H in accordance with a micelle-like pattern of hydrophobicity. On the other hand, residue 314D

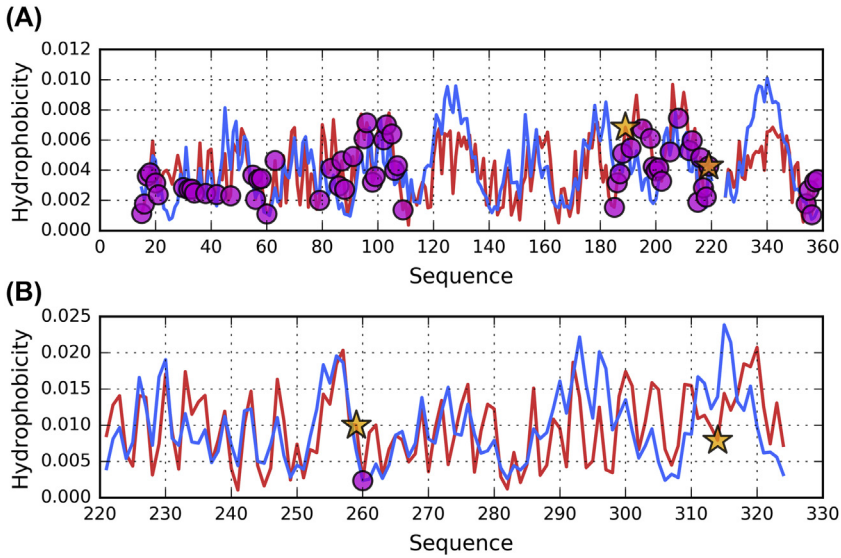


Fig. 8.8 Theoretical (T, blue) and observed (O, red) hydrophobicity distribution profiles for two domains of 1VGM treated as individual units. (A) domain 1 (15−220 + 325−358), (B) domain 2 (221−324) Magenta circles mark residues engaged in P-P interaction, while stars correspond to catalytic residues.

consistently opposes this pattern regardless of the scope of our analysis. This residue therefore carries information by itself, whereas the status of other catalytic residues is determined by interactions between chains forming the dimer (Fig. 8.8).

Conclusions and discussion

In summarizing this part of our analysis, we may conclude that the molecule as a whole differs — somewhat unexpectedly — from a regular micelle, even though environmental factors promote the formation of micellar structures. The protein gains its structural properties in the process of assembly of individual components, each of which folds separately and retains accordance with the theoretical distribution of hydrophobicity (note that the definition of a protein domain includes the requirement for standalone folding). Thus, the information which the protein requires for its biological activity is encoded in the steric arrangement of its constituent domains. The assembly process causes the catalytic active site to diverge from the theoretical distribution.

This conclusion is based on our initial assumption that each protein must fulfill a specific biological role. In the case of enzymes, specific environmental conditions must be produced in order to enable catalysis.

The presented protein is also an example of information encoding through composition. The desired result is achieved by arranging individual components, each of which conforms to the model. When analyzing individual domains on their own, we find that β-structural fragments (β-sheets) conform to the theoretical distribution of hydrophobicity which asserts the presence of a centralized core, and that helical fragments are also consistent with this pattern. If, however, the same secondary folds are analyzed in the context of the protein complex, many of them are found to deviate from the theoretical distribution — consequently, the protein as a whole lacks a clearly defined hydrophobic core.

When attempting to interpret the status of this composite enzyme in light of our introductory hypothesis, we may conclude that — in this case — information is encoded as a specific arrangement of distinct domains, none of which carries much information by itself. The assembly process introduces local deviations with respect to the molecule-wide Gaussian, effectively "imbuing" the protein with information. We may even point to specific places where such information is encoded.

Of course, there many ways in which individual structural units (chains, domains) can assemble into larger structures. The analysis presented in this

chapter aims to shed light on the assembly process, and on the structural properties of the resulting composite proteins.

An example of a protein whose complexity is mainly due to secondary structural variations is the lyase discussed in chapter 7. In addition to classic secondary folds, this protein also contains areas characterized by alternating bands of high and low hydrophobicity.

Another highly complex proteins subjected to FOD analysis is the GroEl chaperonin, which accompanies other proteins at the folding stage. This complex may be regarded as a true "molecular machine" given the complexity of its mechanism of action [4]. Its structure, consisting of 21 chains in three layers with a heptagonal symmetry, undergoes such strong activity-related deformations [5] that its symmetry is effectively erased. The base symmetry is likely required for the complex to revert to its original form once its activity cycle has been completed.

F-actin provides another example of a molecular machine due to its structural and functional complexity. This protein is subject to further analysis in Ref. [6], focusing on local structural patterns and the presence of alternating bands of variable hydrophobicity.

References

[1] Sałapa K, Kalinowska B, Jadczyk T, Roterman I. Measurement of hydrophobicity distribution in proteins – non-redundant protein data bank. Bio-Algorithms and Med-Systems 2012;8. https://doi.org/10.2478/bams-2012-0023.

[2] Parmeggiani A, Krab IM, Okamura S, Nielsen RC, Nyborg J, Nissen P. Structural basis of the action of pulvomycin and GE2270 a on elongation factor tu. Biochemistry 2006; 45(22):6846–57. https://doi.org/10.1021/bi0525122.

[3] Murakami M, Kouyama T. Crystal structures of two isozymes of citrate synthase from sulfolobus tokodaii strain 7. Biochemistry Research International 2016:1–11. https://doi.org/10.1155/2016/7560919.

[4] Banach M, Stąpor K, Roterman I. Chaperonin structure - the large multi-subunit protein complex. International Journal of Molecular Sciences 2009;10(3):844–61. https://doi.org/10.3390/ijms10030844.

[5] Xu Z, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL–GroES–(ADP)7 chaperonin complex. Nature 1997;388(6644):741–50. https://doi.org/10.1038/41944.

[6] Dułak D, Gadzała M, Banach M, Ptak M, Wiśniowski Z, Konieczny L, Roterman I. Filamentous aggregates of Tau proteins fulfil standard amyloid criteria provided by the fuzzy oil drop (FOD) model. International Journal of Molecular Sciences 2018; 19(10):2910. https://doi.org/10.3390/ijms19102910.